

IBM Research Report

An Enterprise Electronic Contract Management System using Dual XML and Secure PDF Documents

Thomas Kwok, Thao Nguyen
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598



Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

An Enterprise Electronic Contract Management System using Dual XML and Secure PDF Documents

Thomas Kwok and Thao Nguyen
IBM Research Division
Thomas J. Watson Research Center
19 Skyline Drive, Hawthorne, NY 10532
kwok@us.ibm.com

Abstract

As most current electronic contract management systems work well for only a few but not all processes in a contract lifecycle because of their inability to handle electronic documents in more than one file format simultaneously. In this paper, we present an efficient and automated enterprise electronic contract management system using both XML and secure PDF documents. The advantages of such dual file formats are to provide the best capabilities and functionalities for contract creation, negotiation, execution and data mining. The structured XML document allows an easy and flexible way to create, negotiate, revise and finalize a contract while the PDF document permits adding signature as watermark on a secure signed contract. A novel conversion method to convert between these two file formats is also described.

1. Introduction

Today, most business between enterprises is conducted under contract. Contracts constitute the binding relationship between an enterprise and its customers, partners or suppliers. Everyday, millions of contracts are created, negotiated, executed and managed via a combination of paper-based manual processes and automatic processes on electronic documents. In most enterprises, contracts in the form of electronic documents can be created using different text editors, such as Microsoft Word. They can also be created by legacy systems and then converted to portable document format (PDF) documents. Some contracts are received in image files from emails sent by customers or partners using scanners. Some other contracts arrive from facsimile machines or in paper documents through mails. As a result, it is very difficult for an electronic contract management system to negotiate, process, execute and data mining all these contracts in either paper documents or electronic documents in various file formats. On one extreme, a structured type document, such as hypertext markup language (HTML) or extensible markup language (XML), is a perfect fit for contract negotiation on-line or over a network (e.g., the Internet), execution and data

mining as it is very easy to parse, extract, and modify document content. However, a structured type document does not provide enough security functions. It is also very difficult to add signing information as watermark on a signed XML contract. On the other extreme, a PDF document, which emphasizes on tight security control, is well-suited for activities such as superimposing signature information as watermark on a signed PDF contract. Nevertheless, the PDF document does not allow negotiation or revision because it only allows line editing. It is also very difficult to carry out data mining in secure PDF documents as it is rather difficult to access, parse and extract data from their content. In between of these two extremes, some text editors, such as Microsoft Word, provide few functions in document security, and tracking modified text or comments. For negotiation and revision, they provide users with limited on-line collaboration capabilities to modify text. However, their negotiation flexibilities are not as good as the structured XML document, and their document securities are not as good or broad as the secure PDF document.

With the advent of Internet technology and electronic commerce, there are growing research activities and implementation efforts on establishing and standardizing an electronic contract management system. Most of the research activities reported focus on contract creation or representation language [1], negotiation and collaboration [2], execution and process [3], signing [4] and data extraction and mining [5]. Currently, the International Association of Contract and Commercial Managers have listed over twenty commercial contract management applications [6]. However, most of these contract management systems work well for only a few but not all processes in a contract lifecycle because of their inability to handle contracts in more than one file format. For example, some systems support PDF documents to enable users to have a tighter security control over their electronic contracts. Other systems provide users with structured XML documents for convenient negotiation and revisions. Further systems employ a text editor with some security features and revision tracking functions for easy contract creation and revision. However, it is very inefficient and cumbersome

to use several systems to create, negotiate, process and data mining an electronic contract during its lifecycle. This approach may also lead to errors or faults, and make it very difficult to automate a contract lifecycle process, which represents a substantial value creation opportunity for the enterprise [7]. Thus, there is a need for an enterprise contract management system that can handle more than one file format type during different process stages of a contract lifecycle. In this paper, we present an efficient and automated enterprise electronic contract management system which can handle electronic documents in dual file formats simultaneously. These two document file formats are the structured XML document and the secure PDF document. The advantages of such dual file formats are to provide the best capabilities and functionalities for electronic contract creation, negotiation, execution and data mining. We also describe a novel method used in this system to switch between these two file formats.

2. The framework

An architectural framework for an enterprise electronic contract management system using dual XML and secure PDF documents is layout in Figure 1. The framework comprises an application server with a local or remote common database, and many Web-based individual user systems. This enterprise Web-based integrated application consists of a number of modules installed in the user's system and the application server. A user system can be a standalone computer system with a local file system. All communications, interactions and connections among user systems, the application server, and the local or remote common database go through one or more secure wired or wireless networks. The application server consists of a XML and PDF converter, the XML and PDF contract generators, a XML and PDF contract exchanger, a generic XML contract negotiation engine, a PDF contract signing engine, the generic index, XML and PDF search engines. Many users, each assigned to perform the same or different tasks on contracts, can simultaneously access their user systems. This user system consists of a contract input module, a contract creation module, a contract negotiation and execution module, and a contract data mining module. The input module enables users to convert those contracts that they have received in other formats into XML and PDF formats by utilizing the XML and PDF converter. If contracts received are originated from scanning paper documents in an optical scanner or a facsimile machine, they are in image formats. The XML and PDF converter will use an optical character recognition (OCR) engine to convert these images into text format. If contracts received are originated from text editors, such as Microsoft Word, as electronic documents, different text editor parsers are used to parse their content

into text format. Then the XML and PDF generators will generate their corresponding XML and secure PDF contracts using their contract templates, XML schema and schema cascading style sheets (CSS), and PDF styles.

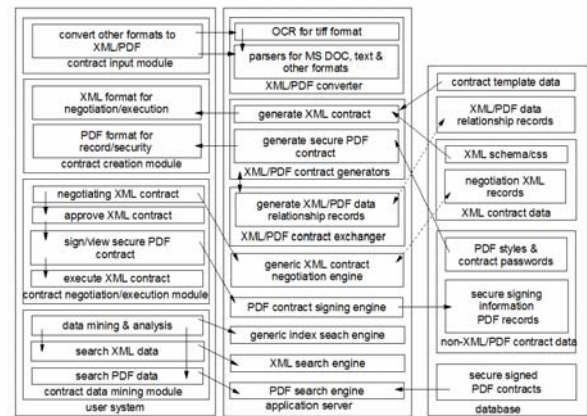


Fig. 1 An architectural framework of an electronic contract creation, negotiation, execution, and data mining system using dual XML and secure PDF documents.

The XML and PDF contract generators are also used by the creation module to create an initial or revised XML contract for negotiation, a final XML contract for execution and process, and a secure PDF contract for viewing, signing and record-keeping. The XML and PDF contract exchanger performs the conversion from a XML contract to a secure PDF contract or vice versa by parsing and separating XML data, non-XML data, such as multipurpose Internet mail extension (MIME) data, and graphic files in graphics interchange format (GIF) or joint photographic experts group (JPEG), and PDF data. It also builds or constructs relationships among separating data as well as generates files that reflect these relationships. In addition to provide all negotiation functions for the XML contract, a generic XML contract negotiation engine also generates and keeps all negotiation XML records. The PDF contract signing engine enables the user to add signing information as watermark on the secure signed PDF contract. The generic index, XML and PDF search engines enable the user in the data mining module to search and analyze previously created contracts.

3. The contract creation module

As indicated in Figure 2, the creation module provides graphical user interfaces (GUIs), graphical panels, templates and scripts to enable a user to create a contract with fewer tedious text input as most terms and conditions can be retrieved from the database. First, a user selects the contract type and parties involved. The user also selects business rules, processes, and signature type associated with the contract. The XML and PDF contract generators will retrieve all these parameters from the contract template data to generate a resultant contract template in accordance with these user-specified

parameters. Then the user can use this contract template to generate a contract with options of reusing existed contract data or inputting new contract data into the contract template.

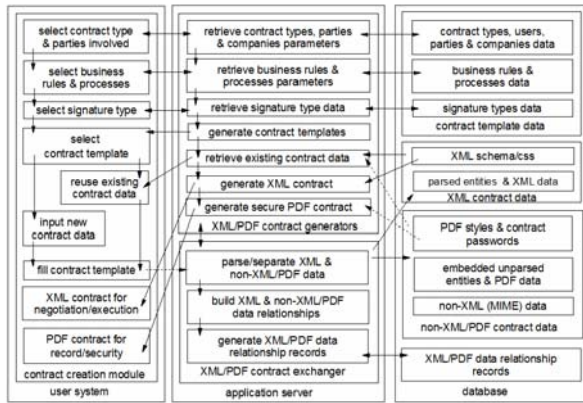


Fig. 2 A schematic diagram illustrating interactions of an electronic contract creation module with the XML/PDF contract generators and exchanger.

In the case of reusing existed contract data, the XML and PDF contract generators will retrieve these contract data from the XML contract data, non-XML data, and PDF contract data. The user can also modify these contract data by changing one or more modifiable fields. Finally, the XML and PDF contract generators will retrieve the XML schema and CSS in the XML contract data to generate a XML contract for negotiation and execution. Similarly, it retrieves PDF styles and contract passwords in the PDF contract data to generate a secure PDF contract for viewing, signing and record-keeping. In the case of inputting new contract data, the XML and PDF contract exchanger will parse and separate XML, non-XML and PDF data from the contract template where the user has filled all empty fields. Then, it will store these data in the corresponding XML contract data, non-XML and PDF contract data. Finally, it will build their data relationships and generate records of these data relationships, as well as store them in the database. The XML and PDF contract exchanger requires these XML and PDF data relationship records in order to convert a XML contract into a secure PDF contract or vice versa.

4. The negotiation and execution module

Figure 3 is a block diagram showing the interactions among the contract negotiation and execution module, the XML and PDF contract generators and exchanger, and the PDF contract signing engine. The negotiation and execution module provides graphical user interfaces (GUIs), graphical panels, templates and scripts to enable a user to select a contract for negotiation, select and connect negotiating parties, negotiate, view, approve and sign the contract. As the user selects a contract and parties involved for negotiation, the XML and PDF contract

generators will retrieve or regenerate a XML contract from its corresponding XML schema and CSS, parsed entities and XML data in the XML contract data. Then the user will be connected to those parties involved by the contract negotiation and execution module to negotiate the contract using negotiation functions provided by any generic XML contract negotiation engine. However, this generic XML negotiation engine is also required to build the negotiation XML records and deposit these records in the XML contract data.

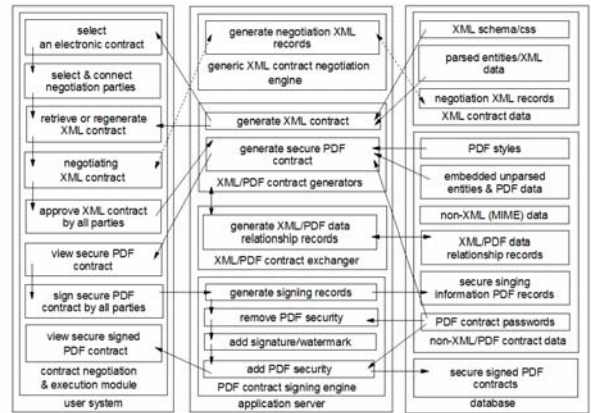


Fig. 3 A block diagram showing interactions among contract negotiation & execution module, XML/PDF contract generators & exchanger, and PDF contract signing engine.

After all pertinent parties involved in the negotiation have approved the final XML contract, the XML and PDF contract generators interacts with the XML and PDF contract exchange to generate a corresponding secure PDF contract for viewing by utilizing XML and PDF data relationship records, as well as retrieving PDF styles and contract passwords. Further modifications on this secure PDF contract are prevented. Then each individual party involved electronically sign this secure PDF contract after viewing it over the network. When a user signs an electronic contract on-line by clicking the "Sign", or any similar acceptance buttons, the PDF contract signing engine captures and gathered the signing information from the signer's computer using the URL get or post method through the network. It also retrieves additional signing information from users and companies data stored in the contract template data. Once all parties involved have electronically signed the contract, the PDF contract signing engine generates the secure signing information PDF records and stores them in the PDF contract data. Then, it retrieves PDF contract passwords from the PDF contract data to remove the document modification security temperately from the PDF contract, and superimposes the signing information as watermark on the PDF contract. Finally, it restores or adds back the document modification security on the signed PDF contract. As a result, a secure signed PDF contract is generated for viewing by users as well as stored in the database for record-keeping purposes.

5. The data mining module

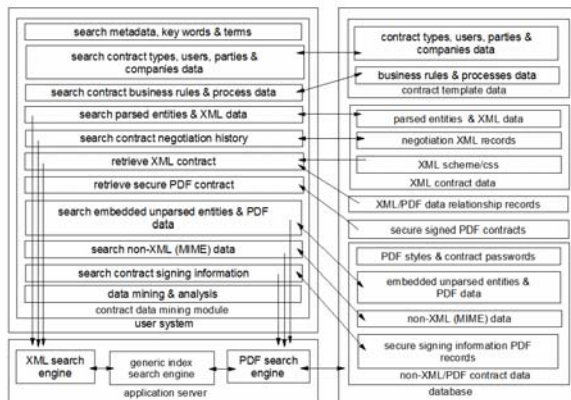


Fig. 4 A block diagram showing the interactions of the contract data mining module in the user system with the application server and the database.

In order to allow the user to carry out data mining and analysis on previously created electronic contracts in both XML and PDF formats, the generic index, XML and PDF search engines have been installed in the application server as illustrated in Figure 4. In addition to metadata and keyword search, the user can search contracts based on contract types, users, parties and company data, as well as business rules and processes data by using a generic index search engine. The user can also search parsed entities and XML data, contract negotiation histories, and XML contracts using the XML search engine. Using the PDF search engine, the user can search embedded unparsed entities and PDF data, non-XML and MIME data, contract signing information and PDF contracts. Moreover, the contract data mining module in the user system provides many other data mining and analysis functions for the user to use by integrating these three search engines together.

6. Conclusion and Discussion

There are many advantages and values in this enterprise electronic contract management system using a XML and a secured PDF document. It provides the best capabilities and functionalities for contract creation, negotiation, execution and data mining. The structured XML document allows an easy and flexible way to create, negotiate, revise and finalize a contract while the PDF document permits adding signature as watermark on a secure signed contract. As the XML security is still being defined and not ready until 2007, the PDF document is used as a secure contract document for final viewing, approval and signing by business and legal teams. The issues of confidentiality, integrity, authenticity and cryptographic non-repudiation of electronic contracts have been discussed in a previous paper [4]. Other advantages include the ability to extract data from previous electronic contracts into the Web-based contract

creation text template for easier and faster contract creation, the capability of capturing and retrieving secure negotiation and electronic signing records. In addition, parsed, unparsed and embedded entities, MIME data and XML elements are stored on database for efficient data search and mining. This approach also enables the automation of a contract lifecycle. As a result, values are stemmed from improved productivity and security, effectively aggregated contract information, accelerated contract lifecycle processes, reduced contractual errors and risk, enabled revenue forecast and profit optimization, as well as better compliance enforcement. Some of features and functions of this system are being implemented in an electronic contract prototype system in IBM research.

10. Acknowledgment

The authors would like to acknowledge the review of the manuscripts, suggestions for improvement and support from G. Pacifici and D. Dias.

11. References

- [1] Y-H Tan and W. Thoen, DocLog: an electronic contract representation language, in Proc. of 35th Annual Hawaii Int'l Conf. on System Sciences, pages 2198-2206, 2002.
- [2] F. Griffel, M. Boger, H. Weinreich, W. Lamersdorf and M. Merz, Electronic contracting with COSMOS – how to establish, negotiate and execute electronic contracts on the Internet, in Proc. of the 2nd Int'l Enterprise Distributed Object Computing Workshop, pages 46-55, 1998.
- [3] T. Kwok and T. Nguyen, A Secure Electronic Contract Management and Process System Automated with Predefined Tasks, in Proc. of IEEE Int'l Conf. on e-Technology, e-Commerce and e-Service, IEEE Computer Society, pages 276-281, 2005.
- [4] T. Kwok and T. Nguyen, An Automatic Electronic Contract Document Signing System in a Secure Environment, in Proc. of 7th IEEE Int'l Conf. on e-Commerce Technology, IEEE Computer Society, pages 497-502, 2005.
- [5] T. Kwok and T. Nguyen, An Automatic Method to Extract Data from an Electronic Contract Composed of a Number of Documents in PDF Format, in Proc. of 8th IEEE Int'l Conf. on e-Commerce Technology, IEEE Computer Society, pages 258-262, 2006.
- [6] International Association of Contract and Commercial Managers, <http://www.iaccm.com>
- [7] W.M. McGovern and L. Lawrence, Contracts and Sales: Cases and Problems, Matthew Bender, 1986.